

# Poursuite d'Objets Multirésolution par Pyramide Irrégulière

Pascal Bertolino et Stéphane Ribas

Laboratoire TIMC-IMAG, Institut Bonniot,

Domaine de la Merci, 38706 La Tronche Cedex, France

Tel : (+33) 476 549 489 - Fax : (+33) 476 549 549 - E-mail : [Pascal.Bertolino@imag.fr](mailto:Pascal.Bertolino@imag.fr)

## RÉSUMÉ

Dans la méthode proposée, une pyramide irrégulière est utilisée pour segmenter les images d'une séquence : une pyramide est construite sur la première image de la séquence, puis sa structure est adaptée d'image en image à l'aide d'un traitement de division-fusion. Ce traitement prend en compte les changements entre deux images successives. De cette façon, c'est la même structure pyramidale qui est utilisée pour traiter plus rapidement toute la séquence. Des critères de stabilité permettent d'obtenir un compromis entre la qualité des résultats et la rapidité. Cela revient à choisir la résolution à laquelle est segmentée l'image. Grâce à la modélisation par graphes, les objets obtenus dans une image peuvent être poursuivis dans les images suivantes.

## ABSTRACT

In the presented method, an irregular pyramid is used to segment the successive frames of an image sequence: a pyramid is built with the first image of the sequence, and then it is updated from image to image, using a split-and-merge process that takes into account the changes occurred between two successive frames. Thus, the same pyramid structure is used along the sequence, speeding up the process. Stability criteria allow to have the required compromise between speed and quality, i.e. to look for image evolution at a particular resolution. Thanks to the graph representation, objects obtained in a given image can be tracked along the rest of the sequence.

## 1 Introduction

Le codage de séquences d'images par segmentation [5] est une alternative souple entre les deux approches suivantes [4] :

- les techniques spatiales ou spectrales qui ne prennent pas en compte le contenu sémantique de l'image,
- les méthodes qui modélisent les objets et requièrent une connaissance a priori de la scène.

La pyramide irrégulière [6] [3] est utilisée pour modéliser et segmenter des images numériques avec la possibilité de prendre en compte l'aspect multirésolution de leur contenu [1].

Cette pyramide est constituée d'un empilement de partitionnements dont la résolution décroît depuis la base (l'image originale) jusqu'à l'apex (dernier niveau) grâce à un traitement de type **croissance en parallèle de régions** sur toute l'image. Dans notre approche, la pyramide n'est construite qu'une fois, puis elle est mise à jour pour tirer parti des similarités entre images consécutives : les différences entre images sont d'abord recherchées à basse résolution puis localement à résolutions croissantes si besoin est, grâce à la structure arborescente. La structure de graphe de chaque niveau permet la poursuite des différents objets sans mise en correspondance. Par la suite, on distingue ainsi le traitement de la première image puis le traitement des images suivantes [2].

## 2 Traitement de la première image

Le traitement de la première image est une phase d'initialisation. La construction de la pyramide irrégulière de cette image suit les phases classiques (figure 1) : tout d'abord, un graphe d'adjacence (base ou niveau 1) est construit à partir de la 4 ou la 8 connexité des pixels de l'image originale. Ensuite, le passage du niveau courant au niveau suivant se découpe de la façon suivante<sup>1</sup>.

**Graphe de similarité** Du graphe d'adjacence est extrait un graphe de similarité à l'aide d'un ou plusieurs critères évalués sur chaque couple de sommets adjacents : par exemple, deux sommets sont similaires si la différence de leur niveau de gris moyen est inférieure à un certain seuil.

**Décimation** Une décimation appliquée sur le graphe de similarité permet de choisir localement le sommet le plus représentatif selon un critère donné. Ces sommets survivants composent le graphe de régions du niveau suivant.

**Rattachement** La décimation est effectuée de façon à ce que chaque sommet non-survivant puisse fusionner au sommet survivant voisin le plus similaire.

**Mise à jour des attributs** Chaque sommet survivant met à jour ses attributs en fonction des siens et des sommets non-survivants qui se rattachent à lui (ses fils). La composante connexe correspondante à chaque sommet est

<sup>1</sup> Au premier niveau, chaque sommet est un pixel. Aux niveaux supérieurs, chaque sommet est un ensemble de pixels connexes (région)

son champ récepteur. A chaque niveau, l'ensemble des champs récepteurs constitue une partition de l'image.

**Graphe d'adjacence** Les adjacences du nouveau niveau sont alors créées par rapport aux adjacences du niveau précédent et aux fusions effectuées.

Le niveau suivant ainsi généré est prêt pour être traité à son tour à l'aide du schéma décrit ci-dessus. Le procédé est itéré jusqu'à ce que le nombre de régions ne décroisse plus sur deux niveaux successifs.

L'ensemble de la structure (les différents niveaux du graphe, les différentes partitions de champs récepteurs et les liens inter-niveaux) est conservé en mémoire pour la suite.

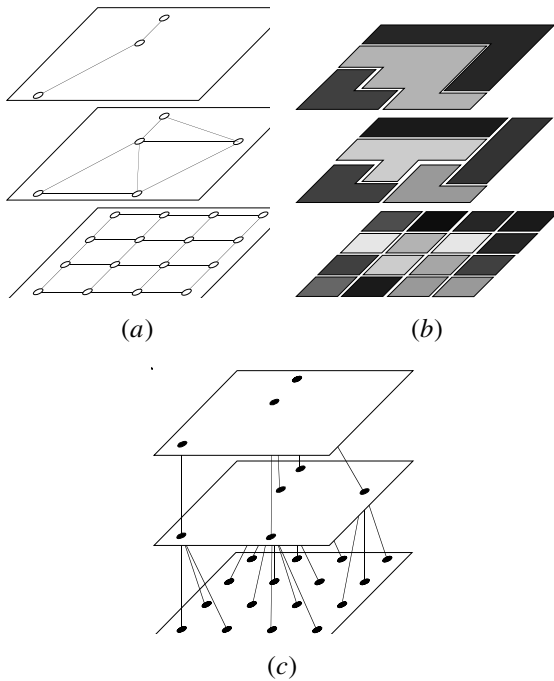


FIG. 1 — Correspondance entre (a) la pyramide de graphes, (b) la pyramide de champs récepteurs et (c) la structure arborescente, sur un exemple d'image  $4 \times 4$

### 3 Traitement des images suivantes

Une fois la pyramide construite sur la première image, elle doit être modifiée pour correspondre au contenu de l'image suivante. Un nombre minimal de modifications doit être appliqué afin qu'une mineure partie de la structure pyramidale soit visitée et mise à jour. Afin de tirer parti de la structure hiérarchique, la correspondance est tout d'abord testée à basse résolution, c'est-à-dire sur le plus petit des graphes (celui représentant l'apex). Si localement la correspondance ne peut être faite, elle est évaluée récursivement au niveau précédent.

#### 3.1 Phase de division

La division s'effectue à partir de l'apex : le partitionnement de l'image  $I_t$  est utilisé pour découper l'image  $I_{t+1}$ . Mais le

contenu des deux images peut être différent et ce partitionnement mal adapté : c'est pourquoi la stabilité de chaque région ainsi obtenue est vérifiée : si les pixels d'une région de  $I_{t+1}$  respectent certains critères de stabilité par rapport à la région correspondante de  $I_t$ , la région est dite *stable* et n'est pas divisée en sous-régions. La sous-arborescence correspondante ne sera pas visitée.

La stabilité est évaluée sur le niveau de gris moyen et l'écart-type moyen des niveaux de gris. De plus, on définit une taille de région minimale au-dessous de laquelle une région n'est pas divisée, quelle qu'elle soit.

Dans le cas où une région n'est pas stable, le traitement de division est itéré sur chacune de ses régions filles. Les régions non stables qui ne peuvent plus être subdivisées (leur taille est inférieure au seuil pré-établi ou il s'agit de pixels) sont appelées régions *modifiées*.

Dans le meilleur des cas, si  $I_t$  et  $I_{t+1}$  sont identiques, seul l'apex de la pyramide et ses champs récepteurs correspondants sont visités, et la pyramide reste intacte. Dans le pire des cas (par exemple passage d'une image uniforme blanche à une image noire), les divisions récursives vont amener à redescendre jusqu'à la base de la pyramide sans trouver aucune stabilité.

#### 3.2 Phase d'intégration

La phase d'intégration permet de prendre en compte les données provenant de l'image  $I_{t+1}$  pour chacune des régions modifiées après la phase de division : chacune d'elles recalcule ses attributs en fonction de la valeur des pixels qu'elle couvre sur  $I_{t+1}$ . Ainsi, elle acquiert des nouvelles valeurs de niveau de gris moyen et d'écart-type de niveau de gris.

Il faut noter que dans le cas où une région modifiée n'appartient pas à la base de la pyramide, ses attributs ne correspondent plus avec ceux de ses filles. Néanmoins, ce problème de non-robustesse n'affecte en rien les traitements et les résultats.

#### 3.3 Phase de fusion

Les parties supérieures de l'arborescence qui ont été divisées doivent à leur tour être mises à jour pour correspondre au contenu de la nouvelle image : la phase de fusion part du niveau le plus bas qui a été modifié, et ne prend en compte que les sommets modifiés ou qui appartiennent au voisinage d'un sommet modifié. Ce traitement permet à la fois d'effectuer la fusion des régions qui ont été déconnectées durant la division, et la reconstruction des parties supérieures des branches de la pyramide en intégrant les données provenant de l'image suivante. La fusion est itérée de niveau en niveau, jusqu'à l'apex. La hauteur de la pyramide peut bien entendu varier sensiblement au cours de la séquence en fonction de la complexité des images.

Hormis le fait que la fusion est effectuée avec un sous-ensemble particulier de sommets (il s'agit cette fois d'une reconstruction partielle de la pyramide), cette croissance de régions procède de la même manière que lors du traitement général énoncé en 2.

## 4 Poursuite d'objets

Avec la structure arborescente de la pyramide et le traitement de division-fusion utilisés, une solution simple à mettre en œuvre permet de suivre les objets d'une segmentation, d'image en image : les  $n$  régions obtenues à l'apex de la pyramide construite sur la première image sont étiquetées (de 1 to  $n$ ). Durant la division, ces étiquettes sont transmises de père en fils. Lors de l'intégration, chaque région modifiée prend l'étiquette NOLABEL, pour indiquer qu'elle ne correspond à aucun objet connu de l'apex. Lors de la fusion, l'étiquette d'un père est mise à jour en fonction de l'étiquette de ses fils, de la façon suivante :

1. Si tous les fils ont l'étiquette  $L_i$ , le père est étiqueté  $L_i$  : l'objet en construction ne comporte que des fragments du même objet.
2. Si tous les fils sont étiquetés  $L_i$  ou NOLABEL, le père est étiqueté  $L_i$  : l'objet en construction comporte des fragments du même objet et des fragments d'un objet inconnu mais aux caractéristiques similaires (puisque la fusion est possible).
3. Si les fils ont différentes étiquettes (par exemple  $L_i$  et  $L_j$ ), le père est étiqueté MULTILABEL. Ce cas peut être dû à une sursegmentation ou à une configuration particulière de la similarité des objets.

Durant la phase de fusion, l'étiquetage est transmis jusqu'à l'apex de la pyramide. Les sommets de l'apex qui conservent une étiquette correspondent à des objets déjà connus sur l'image précédente. Une nouvelle étiquette est attribuée aux sommets étiquetés NOLABEL. Le cas plus complexe des objets MULTILABEL n'est pas traité ici.

## 5 Résultats

Cette méthode a été testée sur des séquences d'images  $256 \times 256$  (figure 2). Le rapport des temps de traitement entre la construction de la pyramide sur la première image et la mise à jour entre images est significatif : la mise à jour est toujours plus rapide, 3 à 8 fois en fonction du paramétrage des critères de stabilité, c'est à dire de la résolution.

Ce paramétrage détermine l'erreur maximale de localisation des changements dans l'image : plus ce critère de stabilité est relâché, plus le nombre de régions modifiées est faible, et plus les traitements s'effectuent à basse résolution (vers l'apex). Ce phénomène peut être observé figure 2 sur la forme irrégulière du cadre mural en haut au milieu des images.

L'optimisation de notre logiciel devrait permettre d'atteindre des traitements qui s'approchent des contraintes du temps réel.

## 6 Conclusion

Nous avons présenté une méthode pour la segmentation des séquences d'images. Grâce à la puissance de la modélisation



FIG. 2 — Segmentation d'une séquence. Colonne 1 : images originales. Colonne 2 : segmentations étiquetées

utilisée, des applications dédiées au codage pour la transmission, à la poursuite en vision par ordinateur ou à l'indexation dans les banques d'images peuvent être envisagées.

Cette méthode nécessitant un recouvrement minimal des objets entre images successives, notre travail se tourne désormais vers l'estimation du mouvement au sein des graphes.

## Références

- [1] P. Bertolino and A. Montanvert. Multiresolution segmentation using the irregular pyramid. In *IEEE ICIP*, pages 257–260, Lausanne, Switzerland, September 16-19 1996.
- [2] J.P. Gambotto. A region-based spatio-temporal segmentation algorithm. In *IEEE proceedings, 11th IAPR, International Conference on Pattern Recognition*, volume 3, pages 189–192, The Hague, The Netherlands, 1992. B.S. Likin and A. Rosenfeld, Academic Press,.
- [3] J.M. Jolion and A. Montanvert. The adapted pyramid : a framework for 2d image analysis. *Computer Vision Graphics and Image Processing*, 55(3) :339–348, May 1992.
- [4] H. Li, A. Lundmark, and R. Forchheimer. Image sequence coding at very low bitrates : A review. *IEEE Transactions on Image Processing*, 3(5) :589–609, september 1994.
- [5] B. Marcotegui. *Segmentation de séquences d'images en vue du codage*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 1996.
- [6] A. Montanvert, P. Meer, and A. Rosenfeld. Hierarchical image analysis using irregular tessellations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4) :307–316, April 1991.